Volume 4 Number 1 (2026)
January – June 2026
Page: 492-504

Journal of Creative Power and Ambition (JCPA)
https://edujavare.com/index.php/jcpa
E-ISSN: 3031-4054
DOI: 10.70610/jcpa.1495

Public Sentiment Analysis of the Burning Sun Scandal Involving Seungri (BIGBANG) on Social Media X Using the Naïve Bayes Method

Purindah Septia Rini[2]

Widya Utama Islamic High School, Indonesia
* E-mail correspondence; purindahseptiarini@gmail.com

Article history Submitted: 2026/02/01; Revised: 2026/03/11; Accepted: 2026/06/23

Abstract

This study aims to analyze public sentiment toward the Burning Sun Scandal involving Seungri by using data from the X social media platform. The dataset consists of 163 tweets collected through a scraping process and processed using text mining techniques, including case folding, cleaning, tokenizing, stopword removal, and stemming. Feature extraction was performed using the TF-IDF method, followed by sentiment classification using the Multinomial Naïve Bayes algorithm with an 80:20 split for training and testing data. The results indicate that public sentiment is predominantly negative. However, the classification model achieved a relatively low performance with an accuracy of 36%. This result is influenced by several factors, such as inconsistent labeling, class imbalance, and the limited size of the dataset. This study demonstrates that the Naïve Bayes method can be applied to sentiment analysis; however, further data processing optimization is required to improve model performance.

Keywords Sentiment Analysis, Naïve Bayes, X Social Media, Text Mining, Burning Sun

INTRODUCTION

The development of information and communication technology has driven the rapid growth of social media as a primary means of expressing public opinion openly and in real time. Social media platforms such as X (formerly Twitter) allow users to express their views, reactions, and emotions on various social, political, and entertainment issues in the form of short texts known as tweets. This makes social media a very rich source of data for understanding public perceptions and sentiments towards a particular phenomenon (Zidan, 2025).

In this context, sentiment analysis is a crucial approach in Natural Language Processing (NLP) for extracting, identifying, and classifying public opinion into specific categories such as positive, negative, and neutral. Sentiment analysis can help systematically and measurably understand public opinion trends toward an issue (Hidayat, 2024). Furthermore, this method has been widely used in various fields, including politics, public services, and even the reputation analysis of public figures.

Sentiment analysis is a branch of Natural Language Processing (NLP) that aims to identify, extract, and classify opinions or emotions contained in a text. In general, sentiment analysis groups text data into positive, negative, or neutral categories based on their meaning (Hidayat, 2024). This approach is crucial for understanding public perception of issues developing on social media. According to Sembiring (2025), sentiment analysis plays a strategic role in processing unstructured data from social media because it can transform subjective opinions into information that can be analyzed quantitatively. In practice, sentiment analysis is widely used in various fields such as marketing, politics, public services, and even analyzing the reputation of public figures.

Methodologically, sentiment analysis can be conducted using two main approaches: lexicon-based and machine learning. The lexicon-based approach uses a dictionary of words weighted by sentiment, while the machine learning approach utilizes classification algorithms to learn patterns from data (Mantika, 2024). In this study, the machine learning approach was used because it offers greater flexibility in handling large and complex data.

Social media platform X is a widely used platform for conveying opinions in real-time in the form of short texts. The main characteristics of this platform are the speed of information dissemination, open access, and high interaction between users. This makes X a highly relevant data source for sentiment analysis research. However, data from social media platform X presents its own challenges, such as the use of informal language, abbreviations, slang, and the presence of noise in the form of emojis, hashtags, and other symbols. Therefore, appropriate preprocessing steps are required to clean and normalize the data before analysis (Zidan, 2025).

Furthermore, social media also reflects the dynamics of public opinion, which can change rapidly as an issue develops. In the context of public scandals such as the Burning Sun Scandal, social media became a primary platform for the public to express support, criticism, and even neutrality toward the figures involved. Text mining is the process of extracting useful information from unstructured text data

through various computational techniques. The stages in text mining include preprocessing, data transformation, feature extraction, and pattern analysis (Syah, 2025).

Stages*preprocessing*is one of the important processes in*text mining*which aims to improve the quality of text data before further analysis. At this stage, the raw text data is cleaned and standardized so that it can be processed more effectively by the system.*preprocessing*starting with*folding case*, namely changing all letters in the text to lower case (*lower case*) so that there is no difference between words that are written in capital letters and lower case letters. Next, do*tokenizing*, namely the process of breaking down text into word units or tokens which will form the basis for text analysis.

The next stage is*stopword removal*, namely removing common words that appear frequently but do not have a significant contribution to the meaning of the analysis, such as conjunctions, prepositions, and pronouns. After that,*stemming*, which is the process of converting affixed words to their base form so that variations of words with the same meaning can be represented in a uniform form. Through these stages, text data becomes cleaner, more structured, and ready for further analysis.

After the process*preprocessing*Once completed, the next stage is feature extraction using the method*Term Frequency–Inverse Document Frequency*(TF-IDF). This method is used to measure the importance of a word in a document relative to the set of documents being analyzed. TF-IDF assigns higher weight to words that appear frequently in one document but rarely in others, thus deeming these words to have more important information. Thus, feature extraction using TF-IDF can improve the quality of text data representation and support more accurate analysis and classification processes (Naraswati et al., 2021).

Naïve Bayes is a probability-based classification algorithm based on Bayes' Theorem. This method assumes that each feature is independent of the others, making probability calculations simpler and more efficient. In general, the basic formula for Naïve Bayes can be explained as follows:

$$P(C \mid X) = \frac{P(X \mid C) \cdot P(C)}{P(X)}$$

Information:

P(C| X) = probability of class C against data X

P(X| C) = probability of data X in class C

P(C) = prior probability of class

P(X) = probability of data

The Naïve Bayes method has several advantages, including fast computation and easy implementation. However, this method also has limitations, namely the assumption of independence between features, which in practice is not always met (Mantika, 2024). In sentiment analysis, the type of Naïve Bayes commonly used is Multinomial Naïve Bayes, because it is suitable for word frequency-based data such as TF-IDF.

Social media X has unique data characteristics, such as unstructured text, the use of informal language, abbreviations, and the presence of noise such as emojis and symbols. Therefore, appropriate text mining and machine learning techniques are needed to process this data to produce meaningful information (Sembiring, 2025). One method widely used in sentiment classification is the Naïve Bayes algorithm due to its simplicity, computational efficiency, and relatively good performance in various previous studies.

Several studies have shown that the Naïve Bayes method can provide fairly accurate classification results in social media-based sentiment analysis. For example, research by Naraswati et al. (2021) showed that this method achieved an accuracy of 87.34% in classifying public sentiment towards government policies on Twitter. Other studies have also shown that Naïve Bayes has competitive performance compared to other methods in classifying public opinion with adequate accuracy (Mantika, 2024). This indicates that the Naïve Bayes algorithm is still relevant for use in sentiment analysis research today.

The Burning Sun scandal involving Seungri of the boy band BIGBANG has become a global issue that has garnered widespread attention from the international community, including social media users. This scandal has not only impacted the South Korean entertainment industry but also sparked intense public debate across various digital platforms. Public reactions to this case reflect how public opinion is formed, developed, and spread rapidly through social media.

However, until now there is still limited research that specifically analyzes public sentiment towards the Burning Sun scandal using a machine learning approach, especially on platform X. Most previous studies have focused more on political issues, public policy, or other social phenomena (Syah, 2025). Therefore, research is needed that specifically examines public perception towards this case to provide a deeper understanding of the dynamics of public opinion towards public figures involved in the scandal.

This study aims to analyze public sentiment towards the Burning Sun scandal involving Seungri by utilizing data from social media X. The method used is Naïve

Bayes with a text mining approach that includes the stages of data collection, preprocessing, feature extraction using TF-IDF, and sentiment classification. Model evaluation is carried out using metrics such as accuracy, precision, recall, and F1-score to measure classification performance comprehensively. This study is expected to contribute to the development of sentiment analysis studies, especially in the context of analyzing public opinion towards public figures in cases of global scandals. In addition, the results of this study can also be a reference for further research in the fields of data mining, social media analytics, and computational social science.

METHODS

This study uses a quantitative research type with a text mining and machine learning approach. The quantitative approach was chosen because this study focuses on processing text data into numerical form which is then analyzed statistically to determine public sentiment patterns. The text mining method is used to extract information from unstructured data in the form of text obtained from social media X. Next, a machine learning approach is applied to classify sentiment using the Naïve Bayes algorithm. This approach is considered effective in identifying public opinion patterns automatically and objectively based on available data.

The data used in this study is secondary data obtained through a scraping process from social media platform X using keywords relevant to the research topic. A total of 163 tweets were collected, which were then used as a dataset in the sentiment analysis process. The research stages in this study follow a systematic flow as shown in Figure 1 (Research Flowchart).

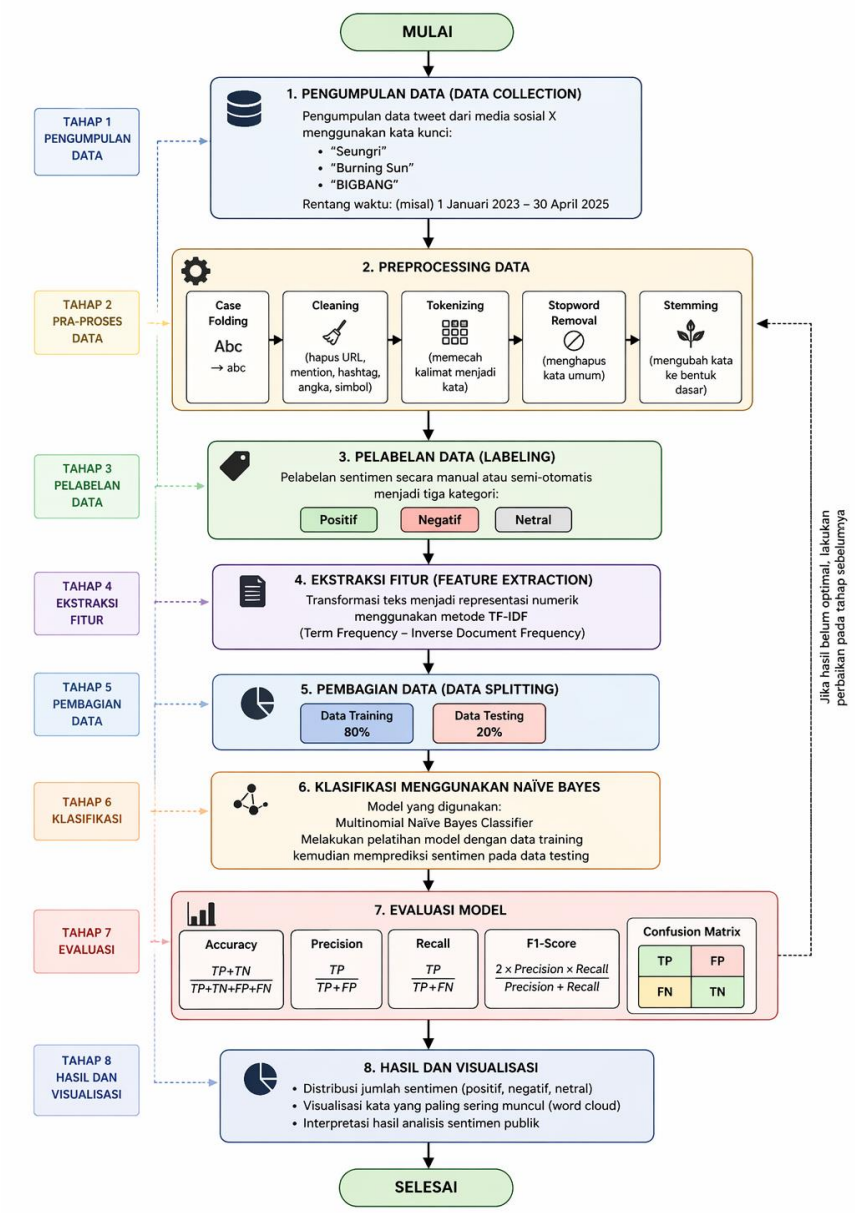

Figure 1. Research Flowchart

This research began with the data collection stage conducted through web scraping techniques on social media X. Data collection used the keywords "Seungri", "Burning Sun", and "BIGBANG" to ensure the data obtained was relevant to the research topic. From this process, 163 tweets were successfully collected which were then used as the research dataset. After the data was collected, a preprocessing stage was carried out to clean and normalize the text data so that it was ready for analysis. This stage aims to eliminate various forms of noise that can affect the analysis results. The preprocessing process includes case folding, which is changing all letters to lowercase; cleaning, which is removing unnecessary URLs, mentions, hashtags, numbers, punctuation, and symbols; tokenizing, which is breaking sentences into

word units; stopword removal, which is removing common words that do not have significant meaning; and stemming, which is changing affixed words to their basic form. The result of this stage is text data that is cleaner, more structured, and ready for the next analysis process.

The next stage is data labeling. At this stage, each preprocessed tweet is manually labeled according to its sentiment into three categories: positive, negative, and neutral. Labeling is performed to generate the training data needed to develop a sentiment classification model. After the data is labeled, feature extraction is performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method. This method transforms text data into a numerical representation by weighting each word based on its frequency of occurrence in the document and its relevance to the overall corpus. Thus, words deemed more important will receive higher weights, helping the model distinguish the characteristics of each sentiment category.

The extracted dataset was then divided into two parts through a data splitting process. Eighty percent of the data was used as training data to build the classification model, while the remaining 20% was used as testing data to evaluate the model's ability to classify previously unstudied data. The sentiment classification process was performed using the Multinomial Naïve Bayes algorithm. This algorithm was chosen due to its strong ability to handle text data and its computational efficiency. The model was trained using the training data to learn the distribution patterns of words in each sentiment category. After the training process was complete, the model was used to predict sentiment on the testing data.

The model's performance was then evaluated using several evaluation metrics, namely accuracy, precision, recall, and F1-score. Additionally, a confusion matrix was used to provide an overview of the number of correct and incorrect predictions in each sentiment category. This evaluation stage aims to determine the level of accuracy and reliability of the model in classifying sentiment. The final stage is presenting the analysis results in the form of data visualization. The visualizations used include sentiment distribution graphs and word clouds that display the dominant words in the dataset. This visual presentation aims to provide a clearer and more easily understood picture of the tendency of public opinion towards the case under study and the sentiment patterns that emerge in conversations on social media X.

FINDINGSAND DISCUSSION

Research result

This study used a dataset obtained from scraping on social media platform X, totaling 163 tweets related to the case under study. The data then underwent preprocessing stages including case folding, cleaning, tokenizing, stopword removal, and stemming to produce cleaner data ready for analysis. After the preprocessing stage, sentiment was labeled into three main categories: positive, negative, and neutral.

Table 1. Dataset Distribution

Total data	163
sentiment	
Negative	64
Neutral	59
negative	18
neutral	12
Positive	7
positive	3

Based on Table 1, the research dataset consists of 163 tweets. The largest sentiment category is Negative, with 64 entries, followed by Neutral, with 59 entries. However, there are inconsistencies in the writing of the labels, namely Negative and negative, Neutral and neutral, and Positive and positive, which are read as different classes by the system. However, based on the results of initial data exploration, it was found that there are inconsistencies in the writing of sentiment labels, such as differences in the use of capital letters for the labels "Negative" and "negative", and "Neutral" and "neutral". This causes the system to identify these labels as different classes, resulting in more than three categories of sentiment classes. This condition results in an unbalanced data distribution and affects the classification results produced by the model.

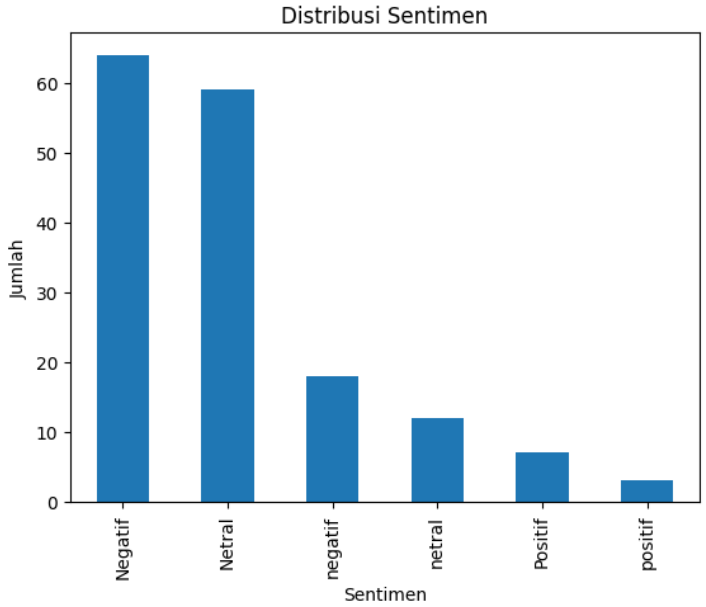

Figure 1. Sentiment Distribution

The distribution of sentiment in the dataset is displayed in the form of a bar graph in Figure 1. Based on this visualization, it is known that the negative sentiment class has the largest number of data, followed by neutral sentiment, while positive sentiment has the smallest number of data. The dominance of negative sentiment indicates that most public opinion towards the analyzed case tends to be critical or unsupportive. The classification process was carried out using the Multinomial Naïve Bayes algorithm with a division of 80% training data and 20% testing data. The model was trained using training data that had gone through a feature extraction process using the TF-IDF method, then tested using testing data to measure classification performance.

Table 2. Classification Report

	precision	**recall**	**f1-score**	**support**
Negative	0.38	0.62	0.47	13
Neutral	0.33	0.33	0.33	12
Positive	0.00	0.00	0.00	1
negative	0.00	0.00	0.00	4
neutral	0.00	0.00	0.00	2
positive	0.00	0.00	0.00	1
accuracy	0.36	33		
Macro avg	0.12	0.16	0.13	33
Weighted average	0.27	0.36	0.31	33

The model testing results showed an accuracy of 36.36%. This value indicates that the model was only able to correctly classify about one-third of the total test data. Furthermore, the evaluation results using the classification report showed that

the precision, recall, and F1-score values for several classes were still relatively low. In fact, several classes scored zero on these metrics, indicating that the model was unable to recognize or predict those classes at all.

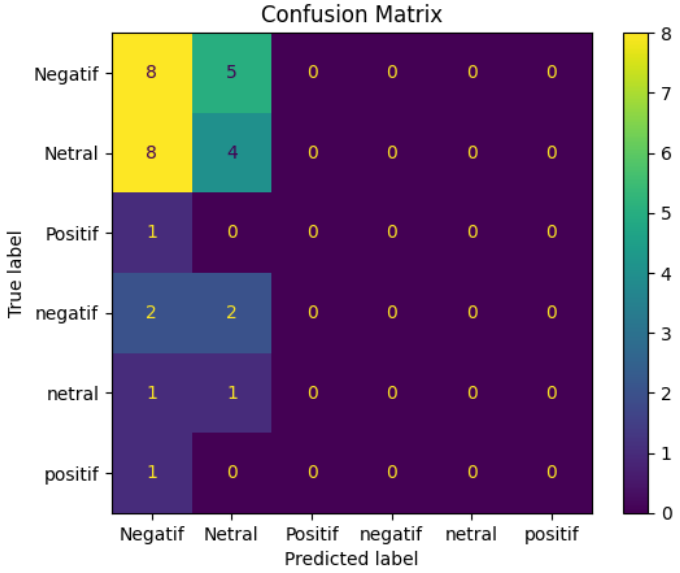

Figure 2. Confusion Matrix

The confusion matrix results shown in Figure 2 show that most of the data tends to be classified into specific classes, especially those with a larger number of data points. This is evident in the greater number of data points predicted as negative compared to other classes. Meanwhile, classes with a smaller number of data points, such as positive sentiment, are poorly identified by the model.

Figure 3. Wordcloud

The word cloud visualization results shown in Figure 3 show that the most dominant words in the dataset include "seungri," "burning," "sun," "case," and "scandal." The appearance of these words reflects the primary focus of public discussion directly related to the research topic. This visualization also shows that the words that appear tend to have negative connotations, which aligns with the dominance of negative sentiment in the dataset.

Discussion

The results showed that the sentiment classification model using the Naive Bayes algorithm produced an accuracy score of 36%. This is relatively low compared to similar studies, which generally range above 70%. This low model performance indicates that several factors influence the model's ability to optimally classify sentiment.

One of the main factors affecting classification results is inconsistency in data labeling. Based on the dataset distribution, it was found that sentiment labels were written in several variations, such as "Negative" and "negative," and "Neutral" and "neutral." The machine learning system treated these labels as distinct classes, increasing the number of classes from three to six. This condition made it difficult for the model to effectively learn the data distribution patterns because the amount of data in each class became increasingly small and scattered. As a result, the model did not have enough data to optimally learn the characteristics of each class. An imbalanced data distribution (imbalanced dataset) is also a significant factor affecting model performance. Research shows that the negative sentiment class has a significantly larger amount of data than the other classes, while the positive class has very little data. This imbalance causes the model to tend to be biased toward the majority class, namely negative sentiment. This is evident in the confusion matrix results, where the majority of data is predicted as the negative class, even though the original labels are different. This condition indicates that the model is more focused on memorizing the dominant class patterns than on understanding the characteristics of each class comprehensively.

Another factor influencing the results was the relatively limited dataset size. With a total of only 163 datasets, this is relatively small for training a machine learning model, especially in the context of sentiment analysis, which involves high linguistic variation. The small dataset limits the model's ability to capture the variety of words and language patterns used by social media users. This results in the model's inability to generalize to new data, resulting in poor performance on the test data. The characteristics of data from social media platform X also present unique challenges in the classification process. Text generated by social media users tends to be informal, containing abbreviations, slang, and mixed language, which are difficult for bag-of-words-based models like TF-IDF to process optimally. Furthermore, the use of irony, sarcasm, and cultural context in the text can also influence sentiment interpretation, increasing the complexity of the classification process.

Nevertheless, the word cloud results show that the most frequently appearing words in the dataset are directly related to the research topic, such as "seungri," "burning," "sun," "case," and "scandal." This indicates that the data used is relevant to the topic being studied. Furthermore, the dominance of words with negative connotations supports the sentiment distribution results, which indicate that public opinion tends to be negative towards the case being analyzed. Compared to previous research, the model performance in this study is still relatively low. This is due to several limitations, such as data quality, dataset size, and suboptimal preprocessing. Previous studies using the Naïve Bayes method generally achieved better results due to the use of larger datasets, consistent labeling, and more complex preprocessing techniques.

CONCLUSION

This study aims to analyze public sentiment towards the Burning Sun scandal involving Seungri by utilizing data from social media platform X and using the Naïve Bayes classification method. Based on the results of the study, it can be concluded that public sentiment obtained from the dataset tends to be dominated by negative sentiment, reflecting the public's critical perception of the case. In terms of model performance, the Multinomial Naïve Bayes algorithm used in this study produced an accuracy value of 36%. This value indicates that the model has not been able to classify sentiment optimally. The evaluation results also show that the precision, recall, and F1-score values are still low in most classes, and some classes are not predicted at all. This indicates that the model has difficulty in recognizing sentiment patterns comprehensively. Several factors that influence the low performance of the model include inconsistencies in data labeling, imbalanced distribution of sentiment classes, a relatively limited number of datasets, and the characteristics of social media data which is unstructured and contains many language variations. These factors cause the model to be unable to learn data patterns optimally.

REFERENCES

Hidayat, R. (2024). Penerapan Naïve Bayes classifier dalam klasifikasi sentimen publik di Twitter terhadap Puan Maharani. Universitas Islam Negeri Sultan Syarif Kasim Riau. https://repository.uin-suska.ac.id/80539/

Mantika, A. M. (2024). Sentiment analysis on Twitter using Naïve Bayes and Logistic Regression for the 2024 presidential election. SANA: Journal of Blockchain, NFTs and Metaverse Technology.

https://journal.mediadigitalpublikasi.com/index.php/sana/article/view/267

Naraswati, N. P. G., Linawati, L., & Putra, I. K. G. D. (2021). Analisis sentimen publik dari Twitter tentang kebijakan penanganan Covid-19 di Indonesia dengan Naive Bayes Classification. Sistemasi: Jurnal Sistem Informasi, 10(1). https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/1179

Sembiring, A. R., & Dewa, C. K. (2025). Sentiment analysis on Indonesian tweets about the 2024 election. Sinkron: Jurnal dan Penelitian Teknik Informatika. https://jurnal.polgan.ac.id/index.php/sinkron/article/view/14481

Syah, F. P., Hasanuddin, T., & Kurniati, N. (2025). Implementasi Naive Bayes untuk analisis sentimen pada data Twitter tentang isu politik di Indonesia. LINIER: Literatur Informatika dan Komputer, 2(3), 302–316. https://jurnal.fikom.umi.ac.id/index.php/LINIER/article/view/3142

Zidan, A. H. F., Handayani, I., & Anggara, A. (2025). Sentiment analysis of the 2024 election using the Naïve Bayes method using data X. Mandiri: Jurnal Akuntansi dan Keuangan, 14(2). https://doi.org/10.35335/mandiri.v14i2.471